

Diploma Thesis—Opponent’s Review

Max Jakob: Mapping the Prague Dependency Treebank Annotation Scheme onto Robust Minimal Recursion Semantics

Description

The thesis describes the conversion of PDT 2.0 tectogrammatical trees into RMRS structures. An algorithm is constructed that associates a RMRS structure to each tectogrammatical node and then composes these structures together. Many tectogrammatical features (including subset of coreference relations) can be represented in RMRS. Valency lexicon is used to distinguish valency and free verbal modifications. The final chapter evaluates the approach, showing promising results. A CD is attached to the thesis with the text itself and documented scripts used to convert the data.

Quality

The thesis is well written, its structure guides the reader through the subject well and its language (English) is clear and without major errors. All the programs on the CD are well documented.

The conversion of the two formalisms has not yet been tried; all the results of the thesis are therefore novel and important. The treatment of various phenomena (e.g. effective parent-child relation) shows the author’s deep understanding of the issue. On the other hand, the author is aware of the limits of his approach, as shown by the lists of lost information. The discussion of evaluation is also quite thorough and can provide a good starting point for future research.

Questions & Remarks

- There are several minor typos in the text. I can provide a detailed list to the author if he is interested. There is also a typographical mistake, a wrong character is systematically used for opening double quotes.
- It is a shame that no quantifier was implemented, at least one “every” would be nice. I did not find a detailed description of the treatment of the negation node #Neg, either.
- Common dependent nodes are processed before members of a coordination on the same level in order to be ready to fill valency slots. However, as I understand it, this only works for one level coordination: in a sentence like *John saw and heard or touched and felt a fitch.*, *John* and *fitch* would not be ready for the verbs.
- The description of the two formalisms involved seems a bit unbalanced. The tectogrammatical layer is presented in quite a detail, while some aspects of MRS (why do adverbs embed verbs?) are left to be sought in the referenced papers. The opposite approach would be more appropriate according to my opinion, since MRS is not well known in our environment, however widely it might have been used elsewhere.

Conclusion

The thesis more than complies with the requirements for Master Thesis at MFF, therefore I recommend to accept it for the defense.

Mgr. Jan Štěpánek, Ph.D.
Karlova Univerzita v Praze
Matematicko-fyzikální fakulta
Ústav formální a aplikované lingvistiky
February the 23rd 2010

